Searching for Prototypical Facial Feedback Signals

Dirk Heylen, Elisabetta Bevacqua, Marion Tellier, Catherine Pelachaud

Searching for Prototypical Facial Feedback Signals

Dirk Heylen[1], Elisabetta Bevacqua[2],
Marion Tellier[2], and Catherine Pelachaud[2]

[1] Human Media Interaction Group, Departement of Computer Science
University of Twente, The Netherlands
[2] IUT de Montreuil
University of Paris8, France

Abstract. Embodied conversational agents should be able to provide feedback on what a human interlocutor is saying. We are compiling a list of facial feedback expressions that signal attention and interest, grounding and attitude. As expressions need to serve many functions at the same time and most of the component signals are ambiguous, it is important to get a better idea of the many to many mappings between displays and functions. We asked people to label several dynamic expressions as a probe into this semantic space. We compare simple signals and combined signals in order to find out whether a combination of signals can have a meaning on its own or not, i. e. the meaning of single signals is different from the meaning attached to the combination of these signals. Results show that in some cases a combination of signals alters the perceived meaning of the backchannel.

Keywords. Feedback, Facial expressions, Interpretation.

1 Introduction

In the context of working on the Sensitive Artificial Listener Agent, a Humaine examplar[1], we are compiling a list of verbal and nonverbal backchannel expressions ([BHPT07], [Hey07]). The goal of the Sensitive Artificial Listener project is to create several talking heads with different personalities that operate as chatbots inviting the human interlocutor to chat and to bring him or her in a particular mood. A particular concern of the project is to have the agent produce appropriate feedback behaviours.

The behaviours displayed by listeners during face-to-face dialogues have several conversational functions. By gazing away or to the speaker a listener signals that he is paying attention and that the communication channels are open. By nodding the listener may acknowledge that he has understood what the speaker wanted to communicate. A raising of the eye-brows may show that the listener thinks something remarkable is being said and by moving the head into a different position the listener may signal that he wants to change roles and say

[1] http://www.emotion-research.net.

something himself. The behaviours that listeners display are relevant to several communication management functions such as contact management, grounding, up-take and turn-taking ([ANA93],[Yng70],[Pog05]). They are not only relevant to the mechanics of the conversation but also to the expressive values: the attitudes and affective parameters that play a role. Attitudes related to a whole range of aspects, including epistemic and propositional attitudes such as believe and disbelieve but also affective evaluations such as liking and disliking ([Cho91]).

Some important characteristics of expressive communicative behaviours are that (a) a behaviour can signal more than one function at the same time, (b) behaviours may serve different functions depending on the context, (c) and behaviours are often complexes composed of a number of behaviours. Moreover, (d) the absence of some behaviour can also be very meaningful.

In this paper we describe a way to gain some further insight in the way certain communicative feedback signals are interpreted. We have used a generate and evaluate procedure where we have asked people to label short movies of the Greta agent displaying a combination of facial expressions. We report here on the second in a series of experiments ([BHPT07]). The aims of these experiments are to get a better understanding of:

- the expressive force of the various behaviours,
- the range and kinds of functions assigned,
- the range of variation in judgements between individuals,
- the nature of the compositional structure (if any) of the expressions.

In this paper, we present the results of the second experiment where we attempted to find some prototypical expressions for several feedback functions and tried to gain insight into the way the various components in the facial expression contribute to its functional interpretation.

A lot has been written about the interpretation of facial expressions. This body of knowledge can be used to generate the appropriate facial expressions for a conversational agent. However, there are many situations for which the literature does not provide an answer. This often happens when we need to generate a facial expression that communicates several meanings from different types of functions: show disagreement and understanding at the same time, for instance. We may find pointers in the literature to expressions for each of the functions separately, but the way they should be combined may not be so easy. In another way, we know that eye brow movements occur a lot in conversations with many different functions. Is there a way in which a distinction should be made between them in terms of the way and the timing of execution or the co-occurrence with other behaviours? In general, listeners make all kinds of comments through their facial expressions, as we will point out in the next section, but the expressions can be subtle.

2 Recognition test

In the previous experiment we found that users could easily determine when a context-free signal conveys a positive or a negative meaning. However, in order to generalise our findings the experiment needs to be performed with more subjects. Moreover as we have tested combinations of signals it occurred to us that we needed to assess the meaning of each single action. Thus, we prepared a second version of the experiment. A first question we wanted to explore with this new test is: is it possible to identify a signal (or a combination of signals) for each meaning? For example, is there a signal more relevant than others for a specific meaning or can a single meaning be expressed through different signals or a combination of signals? We hypothesised that for each meaning, we can find a prototypical signal which could be used later on in the implementation of conversational agents. A second question is: does a combination of signals alter the meaning of backchannel single signals? We hypothesised that in some cases, adding a signal to another could significantly change the perceived meaning. In that case, the independent variable is the combination of signals and the dependent variable is the meaning attributed to each signal by the subjects.

Sixty French subjects were involved in this experiment, the age mean was 20.1 years (range 18-32). They were divided randomly into two groups of thirty: group 1 and group 2.

The test used our 3D agent, Greta [PB03]. Besides the 14 movies used in the previous experiment, Greta displays 7 more movies. Table 1 shows the 21 signals, chosen among those proposed by [AC03,Pog05], that were used to generate the movies. For a more controlled procedure, we decided that participants could not rewind the movie. A list of possible meanings is proposed to the participant who, after each movie and before moving on, can select one meaning according to his/her opinion about which meaning fits that particular backchannel signal best. It is possible to select several meanings for one signal and when none of the meanings seems to fit, participants can just select either "I don't know" or "none" (if they think that there is a meaning but different from the ones proposed). As far as the meanings the subjects have to choose from, we selected: *agree, disagree, accept, refuse, interested, not interested, believe, disbelieve, understand, don't understand, like, dislike.*

1. nod	8. raise eyebrows	15. nod and raise eyebrows
2. smile	9. shake and frown	16. shake, frown and tension[2]
3. shake	10. tilt and frown	17. tilt and raise eybrows
4. frown	11. sad eyebrows	18. tilt and gaze right down
5. tension[2]	12. frown and tension[2]	19. eyes wide open
6. tilt	13. gaze right down	20. raise left eyebrows
7. nod and smile	14. eyes roll up	21. tilt and sad eyebrows

Table 1. Backchannel signals.

Participants were given instructions for the test through a written text in French. They were told that Greta would display back-channel signals as if she was talking to an imaginary speaker. They were asked to evaluate these signals by choosing one or several answers among the available list of meanings. This way we made sure that participants were aware that they were evaluating back-channel signals. The signals were shown once, randomly: a different order for each subject. As the list of possible meanings was too long (12 meanings + none + I don't know), we split it in two, for fear the list might be too long for the subjects to memorise.

2.1 Results

For each meaning, we looked both at the most chosen signals and at the distribution of answers and performed statistical paired t-tests to compare the means of given answers. We especially took a close look at the difference between signals and combinations of signals in order to find out whether adding a signal to another could alter the meaning or not. We present here just the most relevant results. Figure 1 shows about the results the positive meanings.

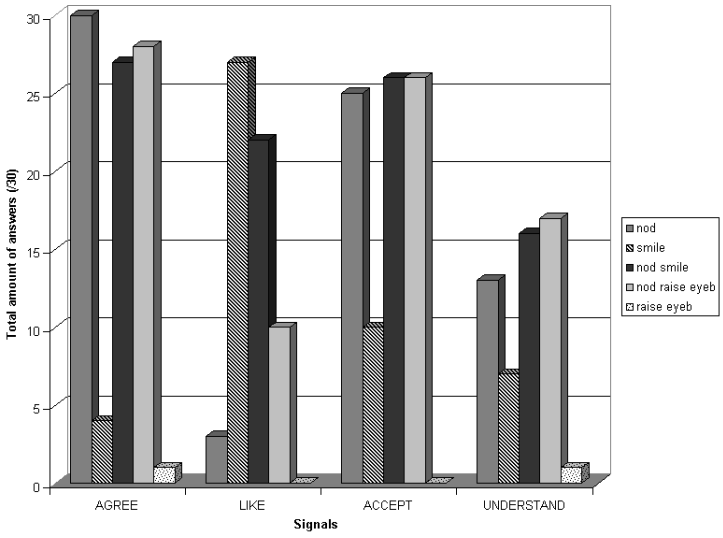

Fig. 1. Signals for positive meanings.

AGREE. When displayed on its own, *nod* proved to be very significant since every subject answered "agree". *Nod and smile* (27 subjects) and *nod and raise*

[2] The action *tension* means tension of the lips.

eyebrows (28 subjects) are also highly considered as backchannel signals of agreement. Difference between the three of them is not significant. When on its own, *smile* (4 subjects) does not mean "agree". For the meaning "agree", difference between the mean of answers for *smile* and the mean of answers for *nod and smile* is highly significant (t=9.761, p<0.0001). We obtained similar results for the meaning of "accept".

LIKE. Two signals convey the meaning "like": *nod and smile* (22 answers) and *smile* (27 answers). The difference between *nod and smile* and *smile* is not significant (t=-1.980, p=0.0573). However, the difference between *nod* (3 subjects) and *nod and smile* is significant (t=-7.077, p<0.0001). This means that the signal *smile* conveys the meaning "like" on its own.

UNDERSTAND. Thirteen subjects associated nod with "understand", 16 paired *nod and smile* with this meaning and 17 found that *nod and raise eyebrows* could mean "understand". There is no statistical difference between *nod* and *nod and smile* (t=-1.795, p=0.0831). There is however a significant difference between *nod* and *nod and raise eyebrows* (t=-2.112, p=0.0434). *Raise eyebrows* on its own does not mean "understand" since only one subject gave that answer.

DISAGREE. The signal *shake* is labelled by every 30 subjects of group 1 as meaning "disagree". The combination of *shake, frown and tension* is also highly recognised as "disagree" (27 subjects). Paired T test shows that there is no significant difference between the two (t=1.795, p=0.0831). The combination of *shake and frown* is also regarded as meaning "disagree" (25 subjects) but it appears that the presence of frown alters the meaning for the difference between the mean of answers for *shake* versus *shake and frown* is significant (t=2.408, p=0.0226). The difference between *shake and frown* and *shake, frown and tension* is not significant (t=-1.439, p=0.1608). In conclusion, *shake* appears as the most relevant signal to mean "disagree", the high and significant difference between *shake, frown and tension* and *frown and tension* (t=10.770, p<0.0001) leaves no doubt about it. We obtained similar results for the meaning of "refuse".

DISLIKE. *Frown and tension* appears as the most relevant combination of signals to represent "dislike" (26 answers). But when *shake* is added to *frown and tension*, it alters the meaning (16 answers). The difference between *frown and tension* and *shake, frown and tension* is significant (t=-3.808, p=0.0007). *Frown* alone is sometimes regarded as meaning "dislike" (by 17 subjects), but it is significantly less relevant that *frown and tension* (t=-3.525, p=0.0014). When displayed on its own, *tension* is also less relevant than the combination *frown and tension*, the difference is significant (t=-4.709, p<0.0001).

DISBELIEVE. Subjects considered that the combination *tilt and frown* means "disbelieve" (21 answers out of 30). It seems that it is the combination of both signals that carries the meaning since *tilt* on its own is regarded as disbelieve by only 8 subjects. Therefore, the difference between *tilt and frown* and *tilt* is significant (t=4.709, p<0.0001). Similarly, *frown* on its own means "disbelieve" for only 6 subjects and thus the difference between *frown* and *tilt and frown* is significant (t=5.385, p<0.0001). Finally, *raise left eyebrow* is also regarded by 21 subjects as "disbelieve".

DON'T UNDERSTAND. *Frown* and *tilt and frown* are both associated to the meaning "don't understand" by 20 subjects. *Tilt* is only given by 4 subjects so that we can infer that *frown* is the most relevant signal of the combination. However, when associated to other signals such as *tension* and/or *shake*, *frown* is less regarded as meaning "don't understand". Difference between *frown* and *frown and tension* is significant (t=2.693, p=0.0117). Similarly, the difference between *frown and tension* and *tension* is significant (t=2.408, p=0.0226), which proves the strong meaning conveyed by the signal *frown*. Apart from the *frown* signal, *raise left eyebrow* appears as relevant to mean "don't understand". It is given by 19 subjects.
NOT INTERESTED. For this meaning, two signals seem to be relevant: *eyes roll up* (20 subjects) and *tilt and gaze* (20 subjects). As far as *tilt and gaze* is concerned, it seems it is the combination of both signals that is meaningful since the difference between *tilt and gaze* and *tilt* (13 answers) is significant (t=-2.971, p=0.0059). Similarly, the difference between *tilt and gaze* and *gaze right down* (13 answers) is also significant (t=-2.971, p=0.0059).

2.2 Discussion

This test provides us with prototypical signals for most of our meanings. For the positive meanings, we have found that "agree" is meant by a *nod*, as well as "accept". To mean "like" a smile appears as the most appropriate signal. A nod associated to a raise of the eyebrows seem to convey "understand" but we have to point out that only 17 subjects out of 30 thought so. As for "interested" and "believe" we will have to test other signals. A combination of *smile and raise eyebrows* could be a possibility for "interested". For the negative meanings, "disagree" and "refuse" are meant by a head shake. Whereas "dislike" is represented by a *frown and tension* of the lips. A *tilt and frown* as well as a *raise of the left eyebrow* mean "disbelieve" for most of our subjects. The best signal to mean "don't understand" seem to be a *frown*. And *tilt and gaze right down* as well as *eyes roll up* are more relevant for the meaning "not interested". It also appeared that a combination of signals could significantly alter the perceived meaning. For instance, *tension* alone and *frown* alone do not mean "dislike", but the combination *frown and tension* does. The combination *tilt and frown* means "disbelieve" whereas *tilt* alone and *frown* alone do not convey this meaning. *Tilt* alone and *gaze right down* alone do not mean "not interested" as significantly as the combination *tilt and gaze*. Conversely the signal *frown* means "don't understand" but when the signal *shake* is added, *frown and shake* significantly looses this meaning. These results contribute to the building up of a library of prototypical backchannel signals.

3 Conclusion

We have presented a perceptual experiment directed to analyse how users interpret context-free backchannel signals displayed by a virtual agent. From our

results we are now able to assign specific signals to most of the meanings proposed in the test and thus begin to define a library of prototypes. Recently, such an experiment has been submitted to subjects of different cultures, in Holland and in Italy. In the future we want to compare the results in order to see if backchannel signals are interpreted in the same way or if they are culture-specific. We also aim at using the set of recognizable signals, defined thanks to this test, in the implementation of a listener model for our conversational agent Greta. Not only the agent will be able to perform such backchannels but, knowing their generic meaning, it will also be able to interpret similar signals emitted by the user. Moreover, this set of recognizable backchannel signals, associated to a set of meanings, opens up further opportunities: we can, for instance, implement virtual agents who display a style of behaviour. For example we can create listeners who appear disbelieving, assertive, not interested and so on and test their effect on users interacting with them.

4 Acknowledgement

Part of this research is supported by the EU FP6 Network of Excellence HUMAINE (IST-2002-2.3.1.6) and by the EU FP6 Integrated Project Callas (FP6-2005-IST-5).

References

[AC03] J. Allwood and L. Cerrato. A study of gestural feedback expressions. In P. Paggio, K. Jokinen, and A. Jonsson, editors, *First Nordic Symposium on Multimodal Communication*, pages 7–22, Copenaghen, September 23-24 2003.

[ANA93] J. Allwood, J. Nivre, and E. Ahlsn. On the semantics and pragmatics of linguistic feedback. *Semantics*, 9(1), 1993.

[BHPT07] E. Bevacqua, D. Heylen, C. Pelachaud, and M. Tellier. Facial feedback signals for ecas. In *In Proceedings of AISB'07: Artificial and Ambient Intelligence*, Newcastle University, Newcastle upon Tyne, UK, April 2007.

[Cho91] N. Chovil. Social determinants of facial displays. *Journal of Nonverbal Behavior*, 15:141–154, 1991.

[Hey07] D. Heylen. Multimodal backchannel generation for conversational agents. In I. van der Sluis, M. Theune, E. Reiter, and E. Krahmer, editors, *Workshop on Multimodal Output Generation*, Aberdeen, Scotland, 2007.

[PB03] C. Pelachaud and M. Bilvi. Computational model of believable conversational agents. In Marc-Philippe Huget, editor, *Communication in Multiagent Systems*, volume 2650 of *Lecture Notes in Computer Science*, pages 300–317. Springer-Verlag, 2003.

[Pog05] I. Poggi. Backchannel: from humans to embodied agents. In *Conversational Informatics for Supporting Social Intelligence and Interaction - Situational and Environmental Information Enforcing Involvement in Conversation workshop in AISB'05*. University of Hertfordshire, Hatfield, England, 2005.

[Yng70] V. Yngve. On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pages 567–577. 1970.